Generalization of a Total Least Squares Problem in Frequency-Domain System Identification

László Balogh and István Kollár, *Fellow, IEEE*

***Abstract*—In this paper, a solution to the frequency-domain system identification of a linear time-invariant system is investigated. A generalization of the total least squares algorithm is shown and analyzed. Some simulation examples on real measured data are given, in order to illustrate the properties of the new method in practice.**

***Index Terms*—Frequency domain, generalized eigenvalues, initial value setting, system identification, total least squares (TLS).**

I. Introduction

MEASUREMENT in a good sense means, e.g., determination of certain properties of a physical system. This is called system identification. Parametric system identification usually concludes in the estimation of unknown parameters in a model [2]–[4]. The estimation of the parameters can be done in many different ways. For the sake of short computing time and numerical simplicity, our goal is usually to cast the problem in the form of a set of linear equations. Because of the distortions and noises in the measurement process, an over-determined set of linear equations is considered. Therefore, an approximation has to be used to make the linear equations compatible. One of these, the total least squares (TLS) method [6], is very effective for frequency-domain system identification. However, in the TLS solution some inherent constraints have to be fulfilled which are sensitive to linear transformations (frequency scaling, etc.). Therefore, it is important to understand what happens during transformations and formulate how the constraints can be transformed. The method described here allows the proper transformation of the constraints, too; the same solution can be found in the transformed space. Therefore, ELiS can be extended to provide TLS solution, independent of frequency scaling (this was not the case until now).

The structure of this paper is the following:

[I] Introduction.

[II] Preliminaries and foundations discusses the notations and assumptions. Furthermore, it contains the basic theorems and statements.

[III] Generalization of the TLS problem contains the theoretical result which is a generalization of the TLS problem.

[IV] Practical use contains verification of the result of the identification and illustrates the practical usage of the new algorithms on real measured data.

Manuscript received May 29, 2001; revised Aubust 23, 2002. This work was supported by the Flemish government (BIL99/18, and GOA-ILiNos) and the Belgian government (IUAP V/22), the Hungarian National Fund for Scientific Research (OTKA: T 033053), and the Hungarian Ministry of Education (FKFP 0074/2001).

L. Balogh and I. Kollár are with the Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest, Hungary (e-mail: lbalogh@mit.bme.hu; kollar@mit.bme.hu).

Digital Object Identifier 10.1109/TIM.2002.808054

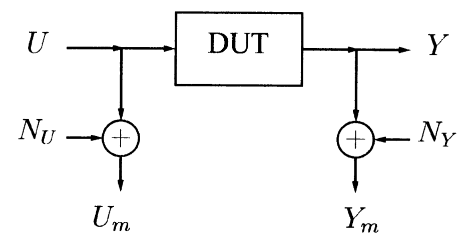

Fig. 1. Measurement setup. DUT is the device under test.

II. Preliminaries and Foundations

In the model, the description of the system with its transfer function is [3]

$$H(j\omega, \mathbf{p}) = \frac{N(j\omega, \mathbf{p})}{D(j\omega, \mathbf{p})} = \frac{\beta_{\mathrm{no}}(j\omega)^{\mathrm{no}} + \cdots + \beta_1(j\omega) + \beta_0}{\alpha_{\mathrm{do}}(j\omega)^{\mathrm{do}} + \cdots + \alpha_1(j\omega) + \alpha_0},$$
$$\mathbf{p}^T = [\alpha_0, \ldots, \alpha_{\mathrm{no}}, \beta_0, \ldots, \beta_{\mathrm{do}}] \tag{1}$$

where ω is the angular frequency, α_i and β_i are the coefficients of the transfer function polynomials, $\mathbf{p}$ is the collection of α_i, β_i, and no and do are the orders of the numerator and the denominator, respectively. A similar expression can be used in the z-domain if $j\omega$ is replaced by $z^{-1} = e^{-j\omega T_s}$, where T_s is the sampling time.

The model of the measurement process can be seen in Fig. 1. The following notations are used:

- U, Y: the exact, but unknown, input, and output;
- N_U, N_Y: additive noises on the input and output, respectively;
- U_m, Y_m: the measured data (Fourier amplitudes at different frequencies).

The following equations describe this stochastic model of the measurement

$$U_m = U + N_U$$
$$Y_m = Y + N_Y. \tag{2}$$

The measured input and output are known at discrete frequencies denoted by $\omega_1, \ldots, \omega_F$. (If time-domain samples are available, the discrete Fourier spectra can be calculated by using the discrete Fourier transform or its fast version, the FFT). It is assumed that the variances of the additive noises are known, and that the noises have zero mean, are uncorrelated over the frequency, and have bounded moments.

If the input/output samples are collected into vectors, we can write

$$\mathbf{U}_m = \mathbf{U} + \mathbf{N}_U$$
$$\mathbf{Y}_m = \mathbf{Y} + \mathbf{N}_Y,$$

where for example $\mathbf{Y}_m$ is

$$\mathbf{Y}_m = [Y_m(j\omega_1) \quad Y_m(j\omega_2) \quad \cdots \quad Y_m(j\omega_F)]^T.$$

Using (1), the model equation is obtained

$$N(j\omega, \mathbf{p})U(j\omega) - D(j\omega, \mathbf{p})Y(j\omega) = 0.$$

This equation is true for every frequency, is linear in $\mathbf{p}$, and can be written in matrix form as

$$\mathbf{Ap} = \mathbf{0} \tag{3}$$

where the rows of the matrix $\mathbf{A}$ belong to the corresponding ω_k. This equation is linear in $\mathbf{p}$, and the elements of $\mathbf{A}$ are

$$\mathbf{A}_{m,n} = U(j\omega_m)(j\omega_m)^{n-1} \quad \text{if } n \leq \text{do} + 1$$
$$\mathbf{A}_{m,n} = -Y(j\omega_m)(j\omega_m)^{n-1-(\text{do}+1)} \quad \text{if } n > \text{do} + 1.$$

Using (2) and (3), the noisy $\mathbf{A}_m$ can be introduced

$$\mathbf{A}_m = \mathbf{A} + \mathbf{N}_\mathbf{A} \tag{4}$$

From the noise assumptions, it follows that

— $\mathbf{N}_\mathbf{A}(j\omega_k), k = 1, \ldots F$ are zero mean, mixing [3], complex random variables;
— $E\{\mathbf{N}_\mathbf{A}\mathbf{N}_\mathbf{A}^T\} = \mathbf{0}$;
— the errors $\mathbf{N}_\mathbf{A}(j\omega_k)$ are independent over the frequency.

For more details, see [3] and [5].

The weighted total least squares

Using (4), the parameter estimation can be formulated as a total least squares problem [5], looking for a solution of

$$\mathbf{A}_m\mathbf{p} = \mathbf{0}$$

where the solution for $\mathbf{A}_m$ may contain errors in all elements. The definition of the weighted TLS problem is the following [6]

$$\min \|\mathbf{W}(\mathbf{A}_m - \hat{\mathbf{A}})\|_F^2 \tag{5}$$

subject to

$$\hat{\mathbf{A}}\mathbf{p} = 0 \quad \text{and} \quad \mathbf{p}^T\mathbf{p} = 1.$$

Here, $\mathbf{W}$ is a left weighting matrix, and F denotes the Frobenius norm. $\hat{\mathbf{A}}$ is an estimation of $\mathbf{A}$. The properties of $\hat{\mathbf{A}}$ and connection with LS can be found in more detail in [6].

Elimination of $\hat{\mathbf{A}}$ in (5) gives the equivalent cost function minimized by the WTLS estimator [5]

$$K = \min \operatorname{trace}((\mathbf{WA}_m\mathbf{p})[\mathbf{p}^T\mathbf{p}]^{-1}(\mathbf{WA}_m\mathbf{p})^T) \tag{6}$$

subject to

$$\mathbf{p}^T\mathbf{p} = 1.$$

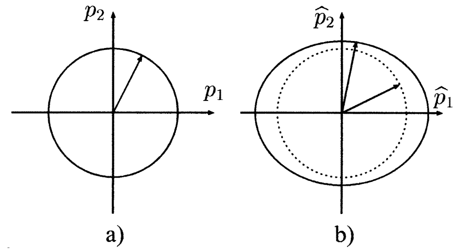

Fig. 2. Original and the transformed space of the parameter vectors.

Transformation of the parameter vector

In many cases, before solution we have to transform the parameter vector into a new base. This can be described as multiplying the parameter vector with a transformation matrix and using the result to continue the estimation algorithm with the vector obtained as the result. The applications of this can be seen in the next section. The transformation of the parameter vector can be written in the following form:

$$\tilde{\mathbf{p}} = \mathbf{Tp}$$

and therefore the $\mathbf{A}_m$ has to be replaced with its transformed version $\mathbf{A}_{mt}$

$$\mathbf{A}_{mt} = \mathbf{A}_m\mathbf{T}$$

Hence the TLS problem should be rephrased, usually like this:

$$\min \|\mathbf{W}(\mathbf{A}_{mt} - \hat{\mathbf{A}}_t)\|_F^2 \tag{7}$$

subject to

$$\hat{\mathbf{A}}_t\mathbf{p}_t = 0 \quad \text{and} \quad \mathbf{p}_t^T\mathbf{p}_t = 1 \tag{8}$$

where $\mathbf{p}_t$ is the variable of the transformed problem. The corresponding cost function is

$$K = \min \operatorname{trace}\left((\mathbf{WA}_{mt}\mathbf{p}_t)\left[\mathbf{p}_t^T\mathbf{p}_t\right]^{-1}(\mathbf{WA}_{mt}\mathbf{p}_t)^T\right).$$

Here, the problem is that the known algorithms cannot account for the fact that by transforming the parameter vector, the constraint $\|\mathbf{p}\|_F^2 = 1$ should be transformed, too. If constraint (8) is used, it is not the original WTLS problem that is solved in the new base.

To illustrate, consider a two-variable parameter vector. Fig. 2(a) shows the original space of the parameter vector with the unit circle as a constraint and the assumed solution of the TLS problem. What will happen if the problem is transformed into a new base? The unit circle is usually transformed into an ellipse. The points of this ellipse are the possible solutions of the original minimization problem. If the generally used algorithm is used, the solution will be searched not on this ellipse, but on the unit circle [see Fig. 2(b)]. It is important to note that in this case after the transformation of $\mathbf{Ap} = \mathbf{0}$, the constraint $(\mathbf{p}^T\mathbf{p})$ is not transformed. In the new algorithm we suggest, the minimization problem is transformed together with the constraint. Hence in the new base the original problem is solved. The new algorithm is discussed in the next section.

III. Generalization of the TLS Problem

The WTLS problem can be generalized in the following way:

$$\min\|\mathbf{W}(\mathbf{A}_m - \hat{\mathbf{A}})\|_F^2$$

subject to

$$\hat{\mathbf{A}}\mathbf{p} = 0 \quad \text{and} \quad \mathbf{p}^T\mathbf{B}^T\mathbf{B}\mathbf{p} = 1. \tag{9}$$

The constraint is a bilinear expression[1]. Hence, the corresponding cost function is

$$K = \min \operatorname{trace}((\mathbf{W}\mathbf{A}_m\mathbf{p})[\mathbf{p}^T\mathbf{B}^T\mathbf{B}\mathbf{p}]^{-1}(\mathbf{W}\mathbf{A}_m\mathbf{p})^T).$$

This problem leads to a generalized eigenvalue problem [3]. Therefore this problem can be solved very effectively with generalized singular value decomposition (GSVD). We may use $\mathrm{GSVD}(\mathbf{WA}, \mathbf{B})$ [7], [6]. The corresponding Matlab program is:

```
[U1,U2,X,S1,S2]= gsvd(W*A,B);
Xi=inv(X');
p=Xi(:,1);
```

This generalization of the constraint allows compensating for the transformation of the parameter vector. If matrix $\mathbf{B}$ is chosen such that

$$\mathbf{B} = \mathbf{T}^{-1}$$

then problem (6) is solved [2]. It means that the solution of the transformed WTLS problem is searched on the transformed unit circle [the ellipse in Fig. 2(b)].

IV. Practical Use

In this section, the applications of the theoretical results are discussed. The focus is on the transformations of the parameter vector.

In practice the transformation of the parameter vector is performed in many cases, although this is not always noticed. We mention here the following occurrences:

— frequency scaling;
— orthogonal polynomial base;
— known subsystem.

Because of the limited length of this article, only the frequency scaling and the orthogonal polynomials will be discussed. The case of the known subsystem can be found in the conference paper [1].

In the following tables, some results obtained by running different algorithms are compared. The normalized difference vector will be used for this purpose. This means that if $\mathbf{r}_1$ and $\mathbf{r}_2$ denote the vectors obtained as two estimation results, the expression of the normalized difference is

$$e = \frac{\|\mathbf{r}_1 - \mathbf{r}_2\|_2}{\|\mathbf{r}_1\|_2}. \tag{10}$$

In our case $\|\mathbf{r}_1\|_2 > 0$.

A. Frequency Scalling

Frequency scaling To avoid the calculation with numbers of different orders of magnitude, which is an ill-conditioned numerical method, the frequencies are first scaled before the estimation algorithms are started [4], [2], [8], [3]. This means that the frequencies are divided by a scale factor which is generally computed in the following way:

$$\omega_{\text{scale}} = \frac{\omega_{\min} + \omega_{\max}}{2}.$$

Essentially, the bandpass spectrum is moved to the radian frequency 1.

Thus, the parameter vector is scaled. Therefore to obtain the final result, the effect of the frequency scaling must be eliminated. This means calculations with

$$\alpha_i \omega_{\text{scale}}^i \quad \text{for } i = 0, \ldots, \text{do}.$$

Similarly, in the case of the denominator,

$$\beta_i \omega_{\text{scale}}^i \quad \text{for } i = 0, \ldots, \text{no}.$$

It can be seen that frequency scaling is equivalent with a transformation of the parameter vector

$$\tilde{\mathbf{p}} = \mathbf{T}\mathbf{p}$$

where

$$\mathbf{T}_{\text{scale}} = \operatorname{diag}\left[\omega_{\text{scale}}^{\text{no}}, \ldots, 1, \omega_{\text{scale}}^{\text{do}}, \ldots, 1\right].$$

Consequently, to solve the original TLS problem, $\mathbf{B}$ must be defined as

$$\mathbf{B} = \mathbf{T}_{\text{scale}}^{-1}.$$

As an illustration of the procedure, the mechanical measurement of a robot arm is presented. The behavior of a flexible robot arm was measured by applying controlled torque to the vertical axis at one end of the arm, and measuring the tangential acceleration of the other end. The excitation signal was a multisine, generated with frequency components at $[1 : 2 : 199]df$, with $df = (500/4096) = 0.125$ Hz, that is, in the frequency range 0.125 Hz–25 Hz. The originally flat multisine was distorted by the nonlinear behavior of the actuator. The odd harmonic frequencies provided that components produced by a squaring nonlinearity would not disturb the identification. The input and output signals were sampled with sampling frequency $f_s = 500$ Hz. Sampling was synchronized to the excitation signal so that 4096 samples were taken from each period. The data records contain 40 960 points, that is, 10 periods were measured. Fig. 3. shows the magnitude of the frequency response at

[1]Note: as a matter of fact (9) can be interpreted that the norm of $\mathbf{p}$ equals one, when the scalar product of vector $\mathbf{x}_1$ and $\mathbf{x}_2$ is defined as $\mathbf{x}_1^T\mathbf{B}^T\mathbf{B}\mathbf{x}_2$.

[2]In the cases of frequency scaling and of orthogonal polynomials, the matrix $\mathbf{T}$ is a square matrix and is invertible. For a known subsytem, $\mathbf{T}$ has to be chosen in another way. See more details in [1].

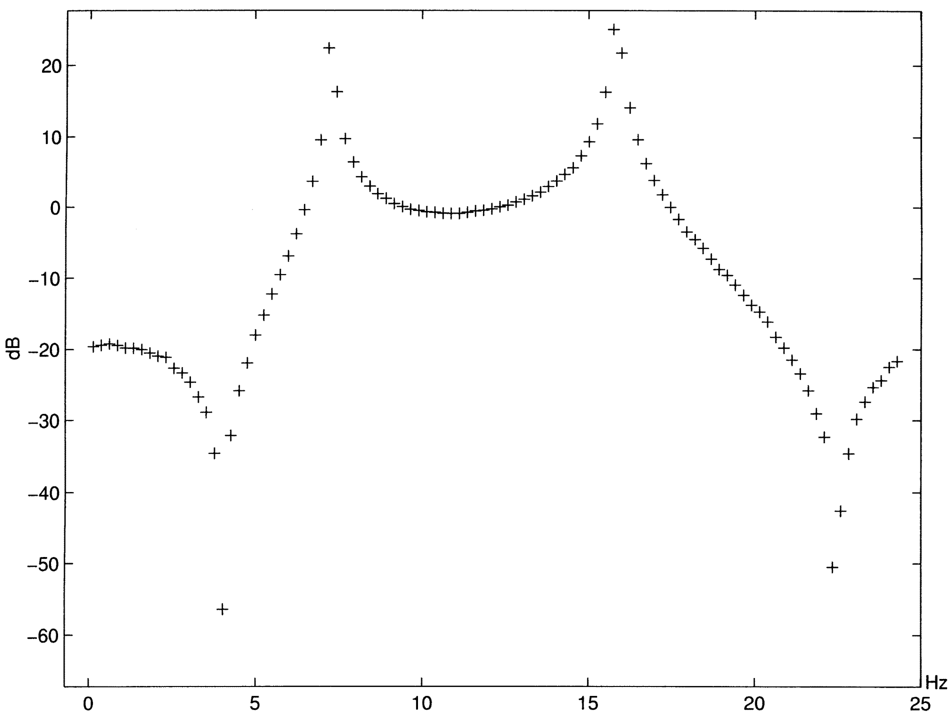

Fig. 3. Measured magnitude of the frequency response of the robot arm.

the measured frequencies. The model is estimated with orders 4/6.

Table I contains the estimation results. The first row of the table is the solution of the original problem. This is the vector to which the others are compared. It can be seen that in the case of the first and the third rows the values of the scaled difference vectors are the same. These vectors are equal. But the parameter vector in the second row differs from the others. The cause is that in this case the bilinear compensation (9) for the frequency scaling was not applied.

B. Orthogonal Polynomials

Orthogonal polynomials are used to enhance the numerical conditioning of the problem. Without details it is important to note that using orthogonal polynomials is equivalent to a base transformation ([8], [3]). If $\tilde{\mathbf{p}}$ denotes a parameter vector in the new base computed with Gram-Schmidt orthogonalization, the transformation can be written as

$$\tilde{\mathbf{p}} = \mathbf{T}_{\mathrm{orth}}\mathbf{p},$$

where $\mathbf{T}_{\mathrm{orth}}$ is the transformation matrix mentioned above. In this case $\mathbf{B}$ has to be set as

$$\mathbf{B} = \mathbf{T}_{\mathrm{orth}}^{-1}.$$

Considering frequency scaling in addition, the value of $\mathbf{B}$ to be used is

$$\mathbf{B} = \mathbf{T}_{\mathrm{orth}}^{-1}\mathbf{T}_{\mathrm{scale}}^{-1}. \tag{11}$$

The same example demonstrated in the previous subsection (robot arm) is used. Equation (11) is applied as the bilinear constraint. Table II contains the estimation results.

TABLE I
RELATIVE DEVIATION OF THE PARAMETER VECTOR

scaling	bilin. comp.	e
no	no	0
yes	no	1.4142
yes	yes	0

TABLE II
RELATIVE DEVIATION OF THE PARAMETER VECTOR

representation	bilin. comp.	e
polynomial	no	0
orthogonal pol.	no	0.7426
orthogonal pol.	yes	0

V. NOVELTIES

In this paper, a generalization of the total least squares problem is discussed, by using a bilinear expression as a constraint of the parameter vector, instead of fixing the norm. Furthermore, an application of this result is shown. It is important because by using the bilinear constraint, the original problem can be solved in the new basis of the parameter vector.

ACKNOWLEDGMENT

The authors would like to thank R. Pintelon for his very useful remarks as well as Haga Kft, Hungary, for their help in performing the measurements and their suggestions in physical modeling.

REFERENCES

[1] L. Balogh and I. Kollár, "Generalization of a total least squares problem in frequency domain system identification," in *Proc. IMTC Conf.*, Budapest, Hungary, May 21–23, 2001, pp. 8–14.

[2] R. Pintelon *et al.*, "Parametric identification of transfer functions in the frequency domain —A survey," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 2245–2260, Nov. 1994.

[3] R. Pintelon and J. Schoukens, *System Identification—A Frequency Domain Approach*. Piscataway, NJ: IEEE Press, 2001.

[4] I. Kollár, Frequency Domain System Identification Toolbox, Gamax, 2001.

[5] R. Pintelon *et al.*, "Analyzes, development, and applications of TLS algorithms in frequency domain system identification," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 4, pp. 983–1004, Oct. 1998.

[6] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem—Computational Aspects and Analysis*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1991.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: The John Hopkins Univ. Press, 1989.

[8] Y. Rolain, R. Pintelon, K. Q. Xu, and H. Vold, "Best conditioned parametric identification of transfer function models in the frequency domain," *IEEE Trans. Automat. Contr.*, pp. 1954–1960, Nov. 1995.

László Balogh was born in Békéscsaba, Hungary, in 1977. He received the M.Sc. degree in electrical engineering from the Budapest University of Technology and Economics, Budapest, Hungary, in 2000. He is currently pursuing the Ph.D.

His research interests include system identification, signal processing, high frequency ground planes, and related applications.

István Kollár (M'87–SM'93–F'97) was born in Budapest, Hungary, in 1954. He received the M.S. degree in electrical engineering from the Technical University of Budapest in 1977, and the "Candidate of Sciences" and "Doctor of the Academy" degrees from the Hungarian Academy of Sciences in 1985 and 1998, respectively.

From 1989 to 1990, he was a Visiting Scientist at the Vrije Universiteit Brussel, Brussels, Belgium. From 1993 to 1995, he was a Fulbright Scholar and Visiting Associate Professor in the Department of Electrical Engineering, Stanford University, Stanford, CA. Currently, he is Professor of electrical engineering at the Budapest University of Technology and Economics. His research interests span the areas of digital and analog signal processing, measurement theory, and system identification. He has published about 70 scientific papers, and is coauthor of the book *Technology of Electrical Measurements* (New York: Wiley, 1993). He authored the Frequency Domain System Identification Toolbox for Matlab. He was Editor of *Periodica Polytechnica Ser. Electrical Engineering* from 1988 to 1991, then Chief Editor of *Periodica Polytechnica* from 1991 to 1997. He was technical cochairman of IMTC 2001.

Dr. Kollár is an active member of EUPAS.